A Survey Paper on Fraud Detection on Event Logs using Constraint Based Association Rule Mining

KamaljitKaur Badwal[1] Anju Ranavadia[2] Rakesh Shah[3]
[1,2,3]Grow More Faculty of Engineering, India

Abstract— ERP is used in most of the businesses to operate it. When integrating many processes like banking, finance, marketing, HR etc. together, it keeps changing itself. Here, there are possibility of data based fraud because transaction are varied and have different sources to process on. There is a approach called association rule mining with process mining can be apply to detect such fraud. Process mining technique mine the fraud in the business process by checking the mismatch between event logs of business process with business process model of a company, which is better known as Standard Operating Procedure (SOP).As the event logs of a company is secured data for them, we will get results on artificial data of a school or a business with customized ERP. Here, the process fraud detection ratio is 96%, but they have considered limited resources. And because the data is not original the result can be processed on likewise criteria. We will try to get it processed with more trust worthy data.

Key words: Process Mining, Event Log, Constraint Based Association Rule Mining, Fraud Detection

I. INTRODUCTION

Enterprise Resource Planning (ERP) has evolved as an integration tool, which is used to integrate all enterprise applications into a data storage center. ERP is easy to access in order to produce high efficiency for the company. However, in its application, business process always changes dynamically. The changes can occur due to several things, such as modification happened in the requirements or a policy in a business process that has been run before [1]. The changes will cause many variations to the original business process. In practice, variations of business process are sometimes executed correctly and incorrectly. The incorrect one may be contains some cheats, called fraud in terms of business process. (2)Fraud in business process can cause the losses for the company. There are many methods to detect the fraud. One of them is process mining which provides several advantages, i.e. conformance checking and control flow analysis in process mining can be used to compare actual data for original business process mode. (3) In networking an event log is a basic resource that helps provide information about network traffic, usage and other conditions.[4] A data mining process may uncover thousands of rules from a given set of data, most of which end up being unrelated or uninteresting to the users. Often, users have a good sense of which direction‖ of mining may lead to interesting patterns and the form‖ of the patterns or rules they would like to find. Thus, a good heuristic is to have the users specify such intuition or expectations as *constraints* to confine the search space. This strategy is known as constraint-based mining.[5] A Transaction processing system (TPS) is a type of information system that collects , Modifies and retrieves the data transaction of an enterprise. e.g. : airline reservation system , electronic transfer of funds , bank account processing system. [6]There are 5 Security issues in ERP (1) Outdated, unsupported software can lead to crashes and integration issues (2) Insufficient reporting capability can lead to external reporting and a loss of data control (3) Technical personnel and providers have access to make large scale changes to program behavior (4) Delayed updates can lead to software vulnerabilities (5) Lack of compliance with security standards

II. LITERATURE SURVEY

[1] Fraud in business process can cause the losses for the company. There are many methods to detect the fraud. One of them is process mining .This paper discusses about fraud detection by using process mining with Fuzzy Association Rule approach. The evaluation of the proposed method uses case study of land management business process at sugar company in ERP system. First step is process to detect deviations which occur in the event logs generated by all activities in business process. Next step, the deviation data is analyzed by fuzzy association rule learning method to produce association rule and the confidence level. From the results of the experiments which have been conducted in two cases [1] conformance checking method in process mining can be used to detect deviation in business process. [2]Fuzzy association rule method can be used to determine fraud based on business process. [3] In addition, the establishment of specific additional rules for handling non-fraud cases can help to improve accuracy as it reduces the occurrence of False Positive in fraud detection.[2] The Paper Proposed improved BM algorithm, BM algorithm is a classical algorithm of single-pattern matching algorithm.BM algorithm using two kinds of heuristic rules: [1] "Good Suffix" shifts [2] "Bad Character" shift The basic idea of the algorithm is that pattern string P and primary string T are aligned on the left side, and pattern string P has a line comparison from right to left with primary string T. Applying the rule of bad characters if the first character on the right side of pattern string P does not equal the corresponding character of primary string T; otherwise apply the rule of good suffixes. The rule of bad characters is that we used in each round of matches when the first mismatch condition.The improved algorithm is applied to the intrusion detection technology, can reduce False Positive Rate and False Negative Rate of intrusion detection system, and improve detection efficiency.[3] In this paper , a framework for global redundancy minimization (GRM). The redundancy is reduced by applying the GRM framework, and classification accuracy has improved significantly for both unsupervised and supervised feature selection algorithms.the effectiveness of the GRM framework, which minimize the redundancy between selected features , thus, the selected features are expected to be more compact and discriminate.[4] In this paper their are three main contributions: (1) A role mining approach for generating transaction profiles from the user activities recorded in the

security log of an ERP system, and for identifying subset relationships amongst such transaction profiles(2) postulated a number of anomalous, possibly fraudulent, activity scenarios which can be detected using the transaction profiles, and identified such anomalies in non-synthetic datasets, and (3) implemented scenarios that identify violations in proper segregation of duties and have detected such violations using the transaction profiles generated in.[5] In this paper, they addressed the problem of discovering a process model from event data of stored in a relational data source, in particular event data of ERP systems. They proposed to discover a model that describes the process as a set of interacting data objects (of a process), each following its own life-cycle, also called artifacts. They validated approach in two case studies using real-life data from ERP systems. In the case studies, the discovered models accurately describe the real executions of the recorded business processes. The case studies also show that the discovered models provide useful insights into the processes and allow users to identify unusual flows of execution.[6] The classification problem is one of the most fundamental problems in the machine learning and data mining literature. text mining techniques need to be designed to effectively manage large numbers of elements with varying frequencies. Almost all the known techniques for classification such as decision trees, rules, Bayes methods, nearest neighbor classifiers, SVM classifiers, and neural networks have been extended to the case of text data. Recently, a considerable amount of emphasis has been placed on linear classifiers such as neural networks and SVM classifiers, with the latter being particularly suited to the characteristics of text data. In re- cent years, the advancement of web and social network technologies have lead to a tremendous interest in the classification of text documents con- taining links or other meta-information. Recent research has shown that the incorporation of linkage information into the classification process can significantly improve the quality of the underlying results [7] In this paper, two metaheuristics are used. The Bat algorithm is better than the traditional particle swarm optimization algorithm, PSO. From the perspective of classification algorithms, Decision Tree is better than Neural Network when they meet imbalanced datasets. The contribution of this paper is a new way of solving imbalanced data problem in data mining by using metaheuristics in the combinatorial choosing of two keyparameters. This method is believed to be found useful for peer data mining users and researchers. The method is generic, that means instead of Bat or PSO other metaheuristics can apply; likewise, the enhanced version ofSMOTE as proposed in this paper can be coupled with other classification methods as well.[8] Data sets from the SAP systems of three different companies operating in diverse industries were used for the evaluation of the designed artifact. It cannot be concluded that the results hold true for other ERP systems and industries but the implemented mining algorithm exploits the structure of accounting entries that is systemindependent and should therefore be generally applicable. The extension to other ERP systems will be covered in future research.[9] This paper presents an evaluation of applying SMOTE to the imbalanced dataset before using SVM and Naive Bayes classifiers. The effectiveness of SMOTE to the datasets depends on how the datasets were processed before applying the oversampling technique and the type of training and testing is also a factor of acquiring precise results. In applying SMOTE, 10 Folds validation provides better findings compared to 70:30 split based on the results

III. Conclusion

So far I have read the above started papers and studied all the required algorithms and techniques…the next stage in my research will be the formation of algorithm steps and analyze association rule mining process for better results.In this survey paper, we have worked on the log analysis technique for security point of view. Our analysis include a technique to detect anomalies, including work flow errors and low performance, by analyzing unstructured system logs. The technique requires neither additional system instrumentation nor any application specific knowledge. moreover a novel technique needs to be enhanced to extract log keywords messages. The limited number of log key types avoids the problem of dimension in the statistic learning procedure. so that weak performance problem highlighted by previous papers can be solved.

References

[1] Kadek Dwi Febriyanti1, Riyanarto Sarno2, Yutika Amelia Effendi3 1Department of Information Technology Management, 2,3Department of Informatics Institut Teknologi Sepuluh Nopember Surabaya, Indonesia,2017
[2] Tao Liu School of Communication and Engineering Xi'an University of Science and Technology Xi'an City, China, Jing Shi School of Communication and Engineering Xi'an University of Science and Technology Xi'an City, China 2016
[3] De Wang, Feiping Nie, and Heng Huang , 2015
[4] Roheena Khan† Information Security Institute Queensland University of Technology, Brisbane 4001, AUSTRALIA, Malcolm Corney Information Security Institute Queensland University of Technology, Brisbane 4001, AUSTRALIA,
[5] Andrew Clark Information Security Institute Queensland University of Technology, Brisbane 4001, AUSTRALIA,George Mohay Information Security Institute Queensland University of Technology, Brisbane 4001, AUSTRALIA ,JUNE 2010
[6] Charu C. Aggarwal IBM T. J. Watson Research Center Yorktown Heights, NY ChengXiang Zhai University of Illinois at Urbana-Champaign Urbana, IL
[7] Xixi Lu, Marijn Nagelkerke, Dennis van de Wiel, and Dirk Fahland , January 2007
[8] Jinyan Li, Simon Fong, Yan Zhuang Department of Computer and Information Science University of Macau Taipa
[9] Michael Werner, Nick Gehrke 2015
[10] Andrew Christian Flores Department of Information and Computer Sciences, Rogelyn I. Icoy Department of Information and Computer Sciences , Christine F.Peña Department of Information and Computer Sciences ,

[11] Ken D. Gorro Department of Information and Computer Sciences, University of San Carlos Cebu City, Philippines

